MPP-2006-43

Maximum Significance at the LHC and Higgs Decays to Muons

Kyle Cranmer[1] and Tilman Plehn[2]

[1] *Goldhaber Fellow, Brookhaven National Laboratory, USA*
[2] *Heisenberg Fellow, Max Planck Institute for Physics, Munich, Germany and School of Physics, University of Edinburgh, Scotland*

(Dated: May 29, 2006)

We present a new way to define and compute the maximum significance achievable for signal and background processes at the LHC, using all available phase space information. As an example, we show that a light Higgs boson produced in weak–boson fusion with a subsequent decay into muons can be extracted from the backgrounds. We generalize our method to include detector effects and reducible backgrounds and describe how it can be incorporated in parton-level event generators.

arXiv:hep-ph/0605268 v2 26 May 2006

Starting soon the prime task for the LHC will be the search for new particles, for example the Standard Model Higgs boson or new particles as suggested by various scenarios for physics beyond the Standard Model. For such searches modern LHC phenomenology has to assess the experimental sensitivity — usually resorting to an *ad hoc* definition of kinematic cuts. This process of designing cuts to isolate signal-enhanced phase space regions (which essentially emulates the traditional experimental practice) is not necessary at the parton level. In this letter we present a new method of computing the expected statistical significance of a hypothesized signal via direct integration of the likelihood ratio. This expected significance is particularly useful as an upper bound: the maximum possible significance for the given signal and background predictions.

To demonstrate the power of this method, we consider the production of the Standard Model Higgs boson at the LHC via weak–boson fusion with a subsequent decay to muons. Weak–boson fusion production of a Higgs boson with a subsequent decay to tau leptons has been firmly established as the main discovery channel for a light Higgs boson [1, 2] in the Standard Model as well in its supersymmetric extension [3]. Observation of the same process with a decay to muons can experimentally confirm Yukawa couplings and their scaling with the masses for non–third–generation fermions.

The expected significance of a search for $H \to \mu\mu$ was estimated for weak–boson fusion [4] and gluon fusion [5] production modes. For a 120 GeV Higgs boson mass, the best kinematic cuts found in Ref. [4] result in a 1.8σ significance. The authors of that analysis note that many observables display additional discriminating power, and suggest that neural networks or other multivariate procedures could enhance the sensitivity. This is a common and unnecessary situation: due to a sub-optimal analysis technique the sensitivity of an experiment cannot be estimated conclusively. Using our new method we indeed find that the maximum possible (target) significance for $H \to \mu\mu$ is much higher, *i.e.* the cut analysis can be greatly improved.

Neyman–Pearson Lemma

Our approach is based on the Neyman–Pearson lemma: the likelihood ratio is the most powerful variable or test statistic for a hypothesis test between a simple (*i.e.* having no free parameters) null hypothesis — background only — and an alternate hypothesis — signal plus background [6]. Maximum power is formally defined as the minimum probability for a Type II error (false negative) for a given probability for a Type I error (false positive). If we assume the signal–plus–background hypothesis is true, the most powerful method has the lowest probability of mistaking the signal for a background fluctuation. This result is commonly used to claim optimality, but these claims can be misleading. The reason is that the probability density function (pdf) of an observable x for a given hypothesis is not experimentally known. Instead, experimentalists typically use a discrete sample of events $\{x_i\}$ to approximately estimate the pdf [7], which invalidates any claims of optimality. In contrast, in phenomenology we can use the parton–level transition amplitude for a given process to exactly compute the pdf over the full phase space.

Two main ingredients are needed to calculate the distribution of the likelihood ratio for the background–only and signal–plus–background hypotheses. First, we have to evaluate identical sets of phase space points for signal and background processes, which is not part of standard Monte Carlo event generators. Secondly, we need to bootstrap the likelihood ratio distribution for one event to the distribution for a fixed luminosity including Poisson fluctuations. Both ingredients are discussed in the next Section. We then consider an example: a light Higgs boson produced via weak–boson fusion and decaying to muons. To achieve a minimum level of realism, we generalize our method to include experimental smearing.

It should be noted that similar techniques are used in experimental analyses. The statistical formalism has been used by the LEP Higgs working group in the Standard Model Higgs search [8]. The smearing of experimental observables is very similar to the methods used at the Tevatron [9, 10]. The major distinction of this work is the combination of the statistical formalism with a parton–level event generator. Note that we do not attempt to identify some powerful discriminating observables, nor do we attempt to compute an observed significance based on experimental data [11]. Instead, we formulate and answer the mathematically well defined question: what is the maximum significance of a new physics signal, *e.g.* a Higgs decaying to muons?

Likelihood Ratio and Discovery Potential

We first limit ourselves to a signal process and its irreducible backgrounds, *i.e.* signal and background processes with identical degrees of freedom in the final state, distinguished by (kinematic) distributions. To compute the expected signal and background rates we integrate the matrix elements squared over the phase space, with or without cuts, using a Monte Carlo integration. This method probes the phase space with random numbers. Ideally, the dimension of the random number vector $\vec{r}$ is given by the number of degrees of freedom in the final–state momenta after all kinematic constraints. The random number vector forms a (minimal) basis for all final-state configurations. We can schematically write

$$\sigma_{\rm tot} = \int_{\rm cuts} dPS\; M_{PS}\; d\sigma_{PS} = \int_{\rm cuts} d\vec{r}\; M(\vec{r})\; d\sigma(\vec{r}) \quad (1)$$

where the phase space boundaries are included in the integral, and the differential cross section $d\sigma(\vec{r})$ includes all phase space factors and the Jacobian for transforming the integration to the random–number basis. The measurement function M can include additional cuts, but we can also use it to extract the event weights as a function of an observable. Because the random numbers parameterize the entire phase space, all possibly available information about the process is included in the array of event weights $(M\, d\sigma)(\vec{r})$.

A cut analysis defines a signal–rich region bounded by upper and lower limits on observables and then counts events in that region. Ultimately, the variable that discriminates between signal and background — the test statistic — is simply the number of events observed in this region. We can predict the expected number of background events b and signal events s, which enables us to optimize the experimental sensitivity by adjusting the cut values. More sophisticated techniques use multivariate algorithms, such as neural networks, to define more complicated signal–like regions, but the test statistic usually remains unchanged. In all of these counting analyses, the likelihood of observing n events assuming the background-only hypothesis is simply given by the Poisson distribution $\text{Pois}(n|b) = e^{-b}\, b^n/n!$.

There are extensions to this number counting, assuming we know the distribution of a discriminating observable x (which may be multi-dimensional). We assume that for the background–only hypothesis H_0 this distribution is $f_b(x)$, while for the signal–plus–background hypothesis H_1 it is $f_{s+b}(x) = [s f_s(x) + b f_b(x)]\,/(s+b)$ assuming no interference. Following the Neyman-Pearson lemma, the most powerful test statistic is the likelihood ratio for the entire experiment's data. The total likelihood for the full–experiment observable $\mathbf{x} = \{x_j\}$ can be factorized into the Poisson likelihood to observe n events, and the product of the individual event's likelihood $f(x_j)$:

$$Q(\mathbf{x}) = \frac{L(\mathbf{x}|H_1)}{L(\mathbf{x}|H_0)} = \frac{\text{Pois}(n|s+b)\; \prod_j^n f_{s+b}(x_j)}{\text{Pois}(n|b)\; \prod_j^n f_b(x_j)}$$
$$q(\mathbf{x}) \equiv \ln Q(\mathbf{x}) = -s + \sum_{j=1}^{n} \ln\left(1 + \frac{s f_s(x_j)}{b f_b(x_j)}\right) \quad (2)$$

We compute the normalized probability distributions $f(x)$ from the parton–level matrix elements. This way construct a log–likelihood ratio map of all possible final–state phase space configurations using the probability distributions $d\sigma(\vec{r})/\sigma_{\rm tot}$ for the signal and background hypotheses:

$$q(\vec{r}) = -\sigma_{{\rm tot},s}\; \mathcal{L}\; + \ln\left(1 + \frac{d\sigma_s(\vec{r})}{d\sigma_b(\vec{r})}\right) \quad (3)$$

$\mathcal{L}$ is the integrated luminosity. To construct the single–event probability distribution $\rho_{1,b}(q)$ we combine the background event weight with the log–likelihood ratio map $q(\vec{r})$ from Eq.(3), which in general is not invertable:

$$\rho_{1,b}(q_0) = \int d\vec{r}\; \frac{d\sigma_b(\vec{r})}{\sigma_{{\rm tot},b}}\; \delta\left(q(\vec{r}) - q_0\right) \quad (4)$$

For multiple events, the distribution of the log–likelihood ratio $\rho_{n,b}$ can be computed by repeated convolutions of the single event distribution. This convolution we can either perform implicitly with approximate Monte Carlo techniques [12], or analytically using a Fourier transform [13].

The expected log–likelihood ratio distribution for a background including Poisson fluctuations in the number of events takes the form $\rho_b(q) = \sum_n \text{Pois}(n|b) \times \rho_{n,b}(q)$. To compute this $\rho_b(q)$ from the single–event likelihood $\rho_{1,b}(q)$ given by Eq.(4) we first Fourier transform all ρ functions into complex–valued functions of the Fourier conjugate of likelihood ratio, *e.g.* $\overline{\rho_{1,b}}(\overline{q})$. The n–event likelihood ratio is now given by $\overline{\rho_{n,b}} = (\overline{\rho_{1,b}})^n$ equivalent to a convolution in q-space. The sum over n in the formula for $\rho_b(q)$ has a simple form in the Fourier domain: $\overline{\rho_b} = \exp[b\;(\overline{\rho_{1,b}}-1)]$. For the signal–plus–background hypothesis we expect s events from the $\rho_{1,s}$ distribution and b events from the $\rho_{1,b}$ distribution. Similar to the above formula we have $\overline{\rho_{s+b}} = \exp[b(\overline{\rho_{1,b}}-1) + s(\overline{\rho_{1,s}}-1)]$. This form we can transform back and obtain the log-likelihood ratio distributions $\rho_b(q)$ and $\rho_{s+b}(q)$.

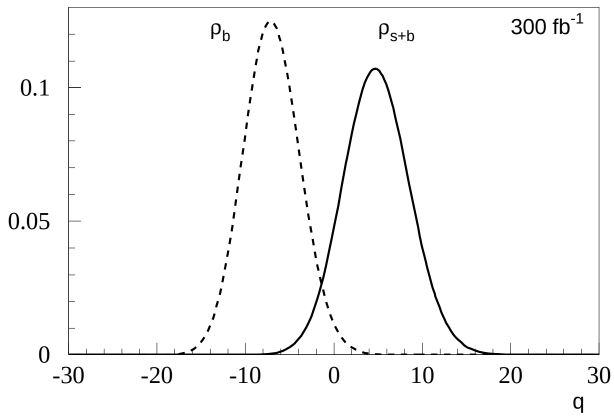

Figure 1: Normalized $\rho_b(q)$ and $\rho_{s+b}(q)$ distributions, corresponding to the full–experiment log–likelihood ratio in Eq.(2). These distributions define the expected significance.

Given a log-likelihood ratio q we can calculate the background-only confidence level, CL_b:

$$\mathrm{CL}_b(q) = \int_q^\infty dq' \; \rho_b(q') \qquad (5)$$

To estimate the discovery potential of a future experiment we assume the signal–plus–background hypothesis to be true and compute CL_b for the median of the signal–plus–background distribution q^*_{s+b}. This expected background confidence level can be converted into an equivalent Gaussian significance $Z\sigma$ by implicitly solving $\mathrm{CL}_b(q^*_{s+b}) = \left(1 - \mathrm{erf}(Z/\sqrt{2}\right)/2$ for Z.

Higgs Decay to Muons

To determine the maximal significance in a strict sense we do not need to include detector effects. However, in our example of weak–boson–fusion $H \to \mu\mu$ the experimental resolution on the invariant mass $m_{\mu\mu}$ is much larger than the Higgs width: about 1.6 GeV for CMS and 2.0 GeV for Atlas [14]. Therefore, we introduce a Gaussian smearing for one of the phase space variables into Eq.(1).

To achieve the smearing, we introduce a new random number r^*_m corresponding to the smeared $m^*_{\mu\mu}$ and integrate over a transfer function from the true $m_{\mu\mu}$ to the smeared $m^*_{\mu\mu}$ by aligning one of the original random numbers r_m with $m_{\mu\mu}$:

$$\sigma_{\mathrm{tot}} = \int d\vec{r}_\perp dr^*_m \int_{-\infty}^{\infty} dr_m \; M(\vec{r}) \; d\sigma(\vec{r}) \; W(r_m, r^*_m) \quad (6)$$

The original random number vector $\vec{r}$ is split into $\vec{r} = \{\vec{r}_\perp, r_m\}$. The transfer function W is a normalized Gaussian giving the likelihood to reconstruct $m^*_{\mu\mu}$ given the true $m_{\mu\mu}$ and the experimental mass resolution. We trivially get back Eq.(1) for $W(r_m, r^*_m) \to \delta(r_m - r^*_m)$.

From Eq.(1) it is obvious how to include an experimental mass resolution: we replace the event weights $(M\, d\sigma)$ by the integral $\left(M \int dr_m \, d\sigma \, W\right)$ and evaluate them over the smeared phase space $\{\vec{r}_\perp, r^*_m\}$. Because the random numbers form a (minimal) basis for all final state configurations there is no 'back door' for the true (infinitely well measured) $m_{\mu\mu}$ to enter the likelihood calculation. A simple approximation incorporating the $m_{\mu\mu}$ mass resolution could be an increased physical Higgs width. It replaces the Gaussian smearing with a Breit–Wigner function; we compare this method with the proper smearing procedure and find that the difference in the final results is small but not negligible.

For all details of the signal and background simulation we refer to Ref. [4]. There, after very basic cuts the signal cross section for a 120 GeV Higgs is 0.22 fb, hidden under 0.33 fb of electroweak Z production and 2.6 fb of QCD Z production, where the Z decays into muons. All other backgrounds combined contribute less than 0.01 fb, which allows us to neglect them.

To probe the likelihood ratio over the full phase space, we relax the cuts for a 120 GeV Standard Model Higgs to mere acceptance cuts. All cross sections are finite, so the cut values have no effect on the likelihood we obtain. Using 2^{20} points we integrate over the final–state phase space projected onto the log–likelihood ratio $q(\vec{r})$ according to Eq.(4). The phase space points used for this integration are defined by the same grid we use for the integration over the signal and background amplitudes described in Eq.(6); this way we can check the total rates to ensure that the likelihood integration covers the entire phase space. For each phase space point we integrate over the true $m_{\mu\mu}$ as shown in Eq.(6), using a proper phase space mapping. Note that this internal integration does not have to use the same grid for signal and background.

The resulting log–likelihood distributions $\rho_b(q)$ and $\rho_{s+b}(q)$ are shown in Fig. 1. From the background pdf we extract the signal significance for an integrated luminosity of 300 fb^{-1} as 3.54σ for CMS and 3.19σ for Atlas. Note that this significance does not include a minijet veto. Following Ref. [4] we can estimate the effect of a minijet veto, which increases the significance to $\sim 4.4\sigma$ for CMS. Combining both experiments the significance even without a minijet veto is 4.77σ.

The most relevant kinematic distribution is the reconstructed Higgs mass $m_{\mu\mu}$. In the upper curves of Fig. 2 we show it for signal and backgrounds without kinematic or likelihood cuts. The signal shows a smeared mass peak, while the backgrounds are flat. To illustrate how the method isolates signal–rich phase space regions, we apply a likelihood ratio cut $q(\vec{r}) > -1.5$. Roughly a third of the signal events survive this cut, and each of the backgrounds are reduced to a rate comparable to the signal. After the likelihood cut the backgrounds show the same kinematic features as the signal, *i.e.* a peak in $m_{\mu\mu}$.

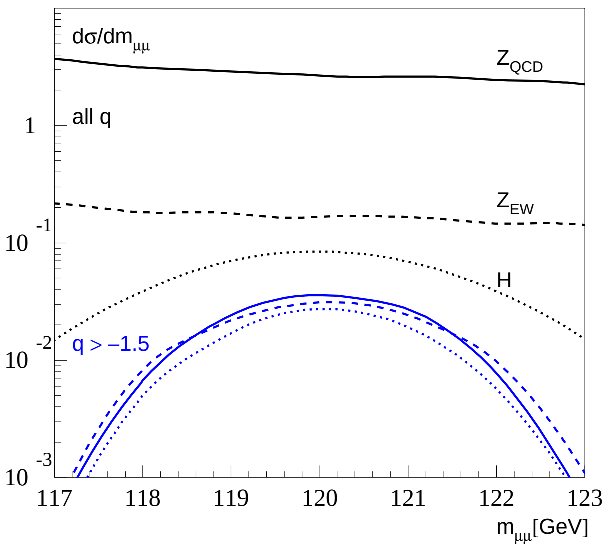

Figure 2: Muon invariant mass distribution for the signal and background with acceptance cuts only (upper curves) and after a cut on the log–likelihood ratio $q(\vec{r}) > -1.5$ (lower curves). The curves illustrate that events with high $q(\vec{r})$ have an increased signal purity and signal–like characteristics.

Detector Effects and Reducible Backgrounds

The procedure for incorporating detector smearing on observables described above is tailored for smearing of a few observables, which are isolated in the phase space integration. Nevertheless, it is possible to generalize the smearing procedure. In essence, a complete detector smearing requires an integration over a fixed set of experimental observables with a nested integration over the remaining degrees of freedom in the phase space. The latter include the unsmeared (true) observables, as shown in Eq.(6), as well as the unobservable longitudinal component of neutrino momenta at a hadron collider or the momentum of particles not passing the acceptance cuts.

We usually include detector effects by smearing all final state four-momenta; however, this can be computationally inefficient. If we instead choose not to smear some of the observables, we must remain vigilant to insure that there is no 'back door' through which four-momentum conservation together with unsmeared observables implicitly evade smearing. We avoid this 'back door' explicitly in Eq.(6) by factorizing the basis of the phase space into orthogonal components r_m and $r_\perp$.

After generalizing our method to smear multiple observables we can now incorporate reducible backgrounds, *i.e.* background whose final–state configurations have more degrees of freedom than the signal. We simply pick a set of observables that is common to all signal and background processes, and marginalize the additional background degrees of freedom. Flavor tagging efficiencies and fake rates can be included in the event weights through $M(\vec{r})$. In these scenarios, the interpretation of the resulting significance is more vague: it is the maximal significance given the specified set of observables and the assumptions in the transfer and measurement functions.

Conclusions

We have described a way to compute the maximally achievable significance for a set of signal and (irreducible) background processes at the parton level. Our method is based on the Neyman–Pearson lemma and can be used to decide if a new physics search at high–energy colliders has a sufficiently large discovery potential to justify a dedicated analysis. It can also be used as a target significance for experimental analyses.

Beyond irreducible backgrounds at the parton level we have shown how to incorporate detector resolution effects for a small number of observables. We have then laid out a recipe for extending our analysis for example to incorporate a fast detector simulation — at the expense of the mathematically strict claim of maximal significance. The next step will be to implement this likelihood computation into a parton–level event generator with a simple and fast simulation of detector effects [15].

Weak–boson–fusion production of a Higgs boson with a subsequent decay to muons is the perfect showcase: it suffers from very low signal rate and from the lack of distinctive signal and background distributions. A cut analysis in Ref. [4] quotes a significance of 1.8σ for 300 fb^{-1} assuming a 1.6 GeV $m_{\mu\mu}$ mass resolution. Applying our method we arrive at a maximum significance of 3.54σ without and $\sim 4.4\sigma$ with a minijet veto. This means that without a luminosity upgrade ATLAS and CMS combined should be able to observe the decay $H \to \mu\mu$.

We would like to thank the Aspen Center of Physics and the theory group of the MPI for Physics for their generous hospitality. Moreover, we would like to thank Dave Rainwater for his help with our example process.

[1] D. L. Rainwater, D. Zeppenfeld and K. Hagiwara, Phys. Rev. D **59**, 014037 (1999), T. Plehn, D. L. Rainwater and D. Zeppenfeld, Phys. Rev. D **61**, 093005 (2000).

[2] for recent overviews see *e.g.*: A. Djouadi, arXiv:hep-ph/0503172; S. Asai *et al.*, Eur. Phys. J. C **32S2**, 19 (2004); V. Büscher and K. Jakobs, Int. J. Mod. Phys. A **20**, 2523 (2005).

[3] T. Plehn, D. L. Rainwater and D. Zeppenfeld, Phys. Lett. B **454**, 297 (1999).

[4] T. Plehn and D. L. Rainwater, Phys. Lett. B **520**, 108 (2001).

[5] T. Han and B. McElrath, Phys. Lett. B **528**, 81 (2002); E. Boos, A. Djouadi and A. Nikitenko, Phys. Lett. B **578**, 384 (2004).

[6] for a proof and corresponding definitions, see *e.g.*: J. Stuart, A. Ord and S. Arnold, *Kendall's Advanced Theory of Statistics, Vol 2A (6th Ed.)* (Oxford University Press, New York, 1994).

[7] K. S. Cranmer, Comput. Phys. Commun. **136**, 198 (2001).

[8] R. Barate *et al.* [LEP Working Group for Higgs boson searches Collaboration], Phys. Lett. B **565**, 61 (2003).

[9] K. Kondo, J. Phys. Soc. Jpn. **57**, 4126 (1998); A. Abulencia *et al.* [CDF Collaboration], arXiv:hep-ex/0512009.

[10] DØnote 5053-CONF (2006), J. C. Estrada Vigil, FERMILAB-THESIS-2001-07.

[11] B. Knuteson and S. Mrenna, arXiv:hep-ph/0602101; B. McElrath *et al.*, in preparation.

[12] T. Junk, Nucl. Instrum. Meth. A **434**, 435 (1999).

[13] for details and a pedagogical discussion of the likelihood ratio see: H. Hu and J. Nielsen, CERN 2000-005, arXiv:physics/9906010.

[14] Atlas Muon TDR CERN/LHCC 97-22. CMS Muon TDR CERN/LHCC 97-32.

[15] K. Cranmer, T. Plehn and J. Reuter, in preparation.